ON06063

93-02035/

OERC CONFERENCE
TORONTO, DECEMBER 1992

BENCHMARKING ONTARIO'S BENCHMARKS

Christine Suurtamm
Peel Board of Education
Meadowvale Secondary School

This study is a result of independent research to clarify the meaning of the term "benchmarks". It is a selective review of how various boards in the province have interpreted or defined the word "benchmark" as well as how those boards use their definition in their assessments of individual students. Research was conducted through interviews with educators, attendance at presentations by developers of specific benchmark programs and through examinations of boards' documents on benchmarks. This research indicates that at the very least, each Board of Education has its own definition of "benchmark". It is this continuum of meanings that is the focus of this paper.

This study began because of my interest in evaluation. As a secondary school mathematics teacher, I see evaluation in mathematics education very overwhelmed by paper and pencil summative tests. However, secondary school mathematics education is beginning to take a fresh look at how students learn mathematics and thus how mathematics is taught. One area of development is in evaluation of students in a mathematics classroom. As forms of mathematics teaching

become more diverse so too must the forms of assessment. As I sought alternative methods of assessment for my classroom I looked to Junior and Primary teachers as well as teachers in other subject areas at the secondary level to give me some insight. I became aware that many of my non-mathematics colleagues were well aware of and quite comfortable with a wide variety of evaluation techniques.

I decided to listen in on the evaluation discussions occurring around me. I discovered that the entire province was in an evaluation discussion and the word I kept hearing in this discussion was "benchmark". I decided to pursue this term "benchmark" to help me increase my understanding of evaluation.

This study is selective. The inquiry into individual board's interpretation of the term "benchmarks" did not cover all boards in Ontario. It was as though I was walking down a corridor and looking in on various board's grappling with benchmarks or actually with evaluation in general. I was not seeking to be evaluative of each board's individual interpretations of evaluation. I was seeking to enhance my understanding by gaining insight from watching others wrestle with the issues of assessment. What I noticed most was the diversity in the approaches to evaluation. It is this diversity that I would like to share.

What is a benchmark? That is often the first question that needs to be answered. The first place that I sought

meaning was out of the context of the educational world. A benchmark is a standard by which something can be measured or judged. It is seen as a reference point. Let's proceed on our journey to look at the definitions of benchmarks in an educational context.

The Ontario Ministry of Education

Burle Summers, the Executive Coordinator for the Ontario Ministry of Education, refers to an education definition of "A benchmark is a standard reference point for assessing student performance on a critial leaning outcome." or "Benchmarks are clear statements of learning outcomes which most students can be expected to achieve at key stages in their education."

As we look at the Provincial Benchmarks it is important to note that these benchmarks are designed to address the accountability of the province for the system of education provided by the Ministry of Education. Benchmarks are to be used to assess both individual students' achievements and educational programs. The Provincial Benchmarks will be used to inform the "auditing" of the province's educational system through provincial reviews. But on an individual student basis, many forms of assessment other than the Provincial Benchmarks should be used. It is the individual boards that address the needs of individual student assessment. It is to those individual boards that I would now like to move.

The Toronto Board of Education

The Toronto Board of Education Benchmarks program began out of the board's concern about accountability to parents. The board felt a responsibility to parents to be able to discuss a child's achievement in relation to system wide standards. (Larter, 1991) Deciding just what those standards would be was a difficult task. Standardized tests were seen as evaluating only a small spectrum of curriculum objectives. Process elements of learning could not be assessed through paper and pencil tests. (Clark, 1990) It was necessary to bridge the gap between testing and learning and it was also thought that teacher observation should be central to the evaluation of students. Bridging the gap between evaluation and learning is a very large task. But that was what was about to be undertaken.

The Toronto Board spent the next several years were spent examining students' performance. A sample of students was chosen to perform specific tasks. These tasks covered many different learning objectives. The tasks themselves were holistic in nature. Analysis, problem solving skills and creativity were incorporated as well as basic skills. These performances were preserved either by videotape or by samples of the students' written work. Committees of teachers were then asked to place these samples of performance on a continuum so that five levels of performance for each task became defined by this range of students' work. Criteria were

then identified for each category of work. "The criteria were thus drawn from the students' performances by the teachers." (Larter, 1991)

The Toronto Benchmarks show a respect for teachers' judgments. Joan Green, Director of Education for the Toronto Board of Education, refers to these benchmarks as providing us with a "window on the learning process". The Benchmarks are a provision of reference points defined by what students are achieving. They reflect the normal learning process. They are realistic goals and yet they are connected to the objectives set out in the Ontario Ministry of Education's guidelines. They constitute a holistic evaluation since any one task takes into account a broad spectrum of learning objectives. Yet, it is a holistic evaluation with very specific criteria to guide the teachers' assessment.

It is important to discuss how these benchmarks are to be used by the classroom teacher. Benchmarks were developed for language arts and mathematics for grades 3, 6 and 8. Further development has occurred to produce benchmarks for the secondary level. The classroom teacher has access to the Benchmark library of videotapes and print. However, the material is not intended to be a set of tests for students in those grades. Each teacher will not be asking all of his or her students to perform each task to be compared to the videotape or samples of work. The samples of work are there to aid the teacher in their observational evaluation. They

will give the teacher points of reference, or benchmarks, so that the teacher can assess where the student is in terms of growth in certain learning objectives.

In attempting to gain a better understanding of the Toronto Benchmarks program I spent a day with an elementary school principal in the Toronto Board. She gave me yet another perspective on the Benchmarks project. For her, Benchmarks are also a way of stimulating good teaching practice. They serve as a catalyst for some excellent professional development. Teachers often work in isolation and using the benchmarks to make classroom activities more visible, more discussable allows for a setting of enhanced growth for the teacher. This principal is using teacher discussion groups as a way of familiarizing the teachers with the library of Toronto Benchmarks. It is through these discussion groups and the examination of the benchmarks that teachers are examining their evaluation and observation practices as well as their own teaching practices.

The Etobicoke Board of Education

The Etobicoke Board of Education has been moving toward using performance based assessment techniques. The goals of using this form of assessment are that the assessment can be cross-curricular, consistent with the various teaching approaches used in Etobicoke as well as being consistent with both board and Ministry of Education guidelines. Dr. Anne

Barens (T.V.O., 1992) also suggests that these assessment techniques will address broad characteristics of the learner that have been set by the board as objectives for the students in the Etobicoke Board of Education. These include creating competent communicators, capable users of technology and lifelong learners.

The Etobicoke board's terminology includes a discussion of assessment targets which are similar to learning objectives. The board also refers to "indicators" which are reference points, benchmarks or levels of achievement. Similar to the Toronto Board, students are given tasks which encompass a variety of curriculum objectives and incorporate different subjects. Students are then evaluated by observation of how they integrate those learning objectives. Problem solving abilities and communication skills within a group would be observed at the same time that specific skills such as writing or calculating are evaluated. Again, we see an integrated, holistic evaluation of the child. This Performance Assessment Program is not yet as highly developed as the Toronto Board's Benchmark Program. It is in its early stages of development and does not reflect the rigor of the Toronto Board but its goals of a holistic evaluation of the student are similar to those of Toronto.

Frontenac Board of Education

The Frontenac Board of Education's current response to the notion of "benchmarks" examines levels of achievement in writing, a difficult area to assess. The Frontenac Board of Education has undertaken a System Review of Writing at the Grade 3, 6, 9, 11 and O.A.C. levels. Each year teachers at these levels are released from their classes for a total of seven and a half days during which time they review samples of students' writing with other teachers teaching at the same level. They are given examples of writing from students from within the board and the teachers are asked to use holistic scoring methods to evaluate the writing and to place the writing into a variety of categories. In using holistic scoring, they are first asked to use a rapid impression marking system and from there to work in groups of three to compare scoring of individual work and to develop reliability checks to this holistic scoring method.

The benchmarks or standard reference points in this project are the rated exemplars of students' work. This evaluated student work is made into a catalog and distributed to all teachers in the board who are involved in evaluating students' writing. Each classroom teacher then uses this catalog of benchmarks as an aid in evaluating the work of his or her own students.

It is important to note that the purpose of these benchmarks is to aid the classroom teacher in evaluation, not

to develop system wide standards. The catalog is to be used within the classroom by the teacher to assess his or her students. It is not intended to be used to compare how individual students are doing in relation to students across the system. This emphasis on aiding individual classroom assessment is an important one. Most of the other benchmark programs discussed in this paper are designed to create system-wide standards or reference points to which the individual can be compared.

North York Board of Education

The "Benchmarks Process" in North York began in the Fall of 1990. This process addresses learning outcomes for reading and writing as well as for mathematics. It is cyclical, with constant testing, analyzing and revising. The Benchmarks Process as outlined in "The Benchmark" (N.Y.B.E., February, 1991) is used to:

> **Identify** an area of the program to be the focus of attention.
>
> **Measure** student learning outcomes using a variety of both formal and informal measures.
>
> **Review** the information gathered to identify areas of strength and areas which need attention.
>
> **Develop** an action plan for addressing each of the four components in the above diagram over the course of a 2-3 year period.
>
> **Implement** the action plan.
>
> **Re-measure, re-examine** the information and the implications and **revisit** whatever steps are necessary.

The project began with student achievement and attitude being examined with standardized tests. These standardized tests were developed by the board and were based on curriculum objectives and guidelines. Programs were then analyzed and modified based on test results. The process involves determining measurable student "outcomes" (or expectations of achievement), measuring those outcomes, reviewing program effectiveness and then developing and implementing plans for change or improvement.

The North York board uses the word "learning outcomes" as direct statements of what students are expected to know and the word "benchmark" to measure or describe student learning at these key points. Teachers are encouraged to use a variety of tools to measure these outcomes. Suggested methods include teacher observation, checklists, student journals, rating scales, self-report inventories as well as tests (situational, norm-referenced, and criterion-referenced). Teachers also could use the results of the board-developed standardized tests to determine whether the students were meeting the benchmarks described in the statements of student outcomes and thus the benchmarks become an aspect of the teacher's evaluation of the student.

Veronica Lacey, Director of Education for North York states in "Student Outcomes in Reading and Writing"(1991) that:

> The outcome statements present a clear picture of North York expectations for all students in reading

> and writing. Together, teachers and parents can use the outcome statements in conjunction with samples of student work and various aassessment instruments to discuss individual student growth.

These benchmarks attempt to address the dilemma of evaluating a wide variety of assessment techniques including standardized testing. The board is responsible for developing the standardized test. It is hoped that the board also sees a responsibility in developing the teachers' expertise in using the other suggested methods of assessment.

The Ottawa Board of Education Benchmarks

The Ottawa Board of Education is also concerned with accountability and system wide standards. The Ottawa Board of Education began a Learning Profiles Project in 1988 (Winks, 1990). This learning profiles project was made up of four committees of which the Benchmarks Committee was one. The purpose of the Benchmarks Committee was to "determine the expectations for social/emotional and academic achievements at the end of each division: grade 3, 6, 9 and 12."(Winks, 1990). At first the benchmarks were determined by looking at Ministry and Board documents. Later these benchmarks were reexamined by looking at students and how they fit into the benchmarks that were determined.

The secondary panel and the elementary panel seem to be taking two different approaches. At the secondary level there has been system wide assessment for quite some time. In the secondary panel, teacher teams have been writing exemplary

tests which are used across the system. There have also been standard course outlines. In the T.V.O. Teleconference on Benchmarks, David Hunter, Coordinator of System Wide Student Evaluation for the Ottawa Board of Education, suggests that the benchmark program at the secondary level takes the position that evaluation should drive curriculum. He suggests that accountability creates seriousness on the teachers part. (T.V.O., 1992) The purpose of boardwide standards of assessment are to monitor and facilitate student learning and to see what is happening in the schools.

A common course curriculum for all secondary level courses and benchmark "tests" used by the Ottawa board are completely generated by teachers in the field and then are distributed for use across the board. This concept of benchmarking seems more closely related to standardization rather than establishing standards. In the T.V.O. presentation David Hunter gave an example of the Grade Nine General Level Mathematics Standards. Each school in the board uses the same mark breakdown for this course. A unit on area and perimeter was used as an example. The common evaluation for this unit consists of 20% of the mark for project, 20% for notebook evaluation, 20% for behavior and social skills and 40% for a final written test. No evidence was given to indicate that all learning objectives are being incorporated in the assessment of the student or the program. It also was not clear whether secondary school teachers' expertise at

using alternative assessment techniques was being enhanced by using these standards at the secondary level.

The Learning Profiles Project is more apparent at the elementary level. There seemed to be three stages of development in this project. The first stage was the development of achievement targets. The next stage was the establishment of levels of achievement of those targets. These levels are referred to as benchmarks. And the third stage was the writing of descriptions or collecting of examples of student work to illustrate those levels. The achievement targets were to be assessed within the context of other work. For example an achievement target might be "spell correctly most high-frequency words". This achievement target would not be assessed by a spelling test but rather in context by a student's spelling within writing assignments.

Thus benchmarks become the levels of achievement in these achievement targets and exemplars or descriptors are given to teachers to show or explain the benchmark. Most of the achievement targets are related to basic skills such as knowing multiplication facts or spelling words. Other learning objectives do not appear in the achievement targets. The Ottawa Board document (1991) states that

> In addition to the achievement targets listed in this booklet, there are essential skills that must be developed and important attitudes that must be fostered for living and learning. These skills and attitudes cross traditional subject area boundaries.

In comparison with the Toronto Board, the Ottawa Board is

focussing on basic skills rather than evaluating the child in a more holistic manner. Evaluation is being carried out in those learning objectives that are easily measured and quantifiable.

The Ottawa Board has developed and is in the process of piloting "Track 21", a computerization of benchmarks. This board is using technology to keep a profile on each student. The plan is for each teacher to have a computer at his/her desk to be used for administrative duties as well as using this tracking program. The program displays each students' name as well as a list of the achievement targets in a chart form. The teacher places a code in each square of the chart to indicate at what benchmark (level) the student is with respect to that achievement target. From a technological standpoint, the entire system is quite intriguing. If the teacher wants further explanation of the achievement target or needs a sample of work at that level then that also is a computer option. There is a great deal of emphasis on this tracking program so that student profiles can be maintained and are accessible as the child moves on through school. Benchmarks seems to have created a heavy emphasis on better reporting of assessment. Accountability seems to play a key role. This system could play a key role in developing teacher expertise at assessment of achievement. It is hoped that the emphasis on tracking the student doesn't deter from the quality of evaluation.

Conclusion

One word "benchmark" has opened up a discussion of evaluation in school boards across Ontario. There may be varied definitions and interpretations given to the term "benchmarks", yet each time a board attempted to define the term for themselves, they had to open the discussion on evaluation. This discussion has made the public, teachers parents and students more aware of the intricacies of trying to define what and how we evaluate. By looking at these few boards, we see that there are a myriad of responses to the term "benchmark". But there are many similarities as well. I would like to look at the common ground that each board addresses before we look at the disparities of the different approaches.

One similarity is that each board was concerned with is accountability. This accountability refers to the individual student's program as well as the system-wide program of education. In most boards the emphasis was to define what can be expected of students in that board at various grade levels.

Another similarity was that each board had a set of learning objectives as well as levels at which those learning objectives could be met. The set of learning objectives may have been called by a different name by each board. In Etobicoke they were called "assessment targets", in North York "learning outcomes" and in Ottawa, "achievement targets". The levels of achievement were called "benchmarks" by many boards

as well as being called "indicators" or "standard reference points". The similarity in the terminology created a similarity in structure. But it is the filling in of that structure that has many differences.

It is important to examine the different ways that the boards move within the common ground of concern with accountability and defininition of learning objectives and their evaluation. Evaluation in education is about values. What we evaluate and how we evaluate display what we as educators value. When we discuss evaluation, we are discussing what really counts in schools. Educators should be able to articulate their objectives, how they are going to achieve those objectives and how they will evaluate the extent to which those objectives have been met. That is what accountability is about.

In examining the accountability of a "benchmark" system we need to look at the articulation and choice of learning objectives as well as the forms of assessing whether those objectives are met. We could, in fact, set up a benchmark system for evaluation of benchmark systems. What would that have to look like? From my perspective, I would think that if a board were to be benchmarking its own benchmarks, that board would have to look at the two concerns, articulation of learning objectives and the evaluation of these objectives as two of their main achievement targets. If we look at these achievement targets, then we also need to look at the possible

levels of achievement or benchmarks for each. Let's look at what criteria we might choose there.

We could consider our first achievement target for a board's evaluation system as being the articulation of its learning objectives. There is a wide variety of choice in the learning objectives. They could be the acquiring of basic skills, learning to integrate many disciplines, being cooperative or being a good problem solver. The differences between the boards fall into place when we look at just what should be included in these assessment targets or learning outcomes. Some boards, such as Toronto and Etobicoke, try to incorporate a wide range of the learning outcomes set by board and Ministry guidelines while others such as Ottawa seem to place an emphasis on those learning outcomes that are more easily measured by standard reference points. It would seem that the most difficult task is to attempt to measure those learning outcomes that we have always had difficulty measuring. We should not shy away from this difficulty. We should not just focus on those objectives that are easily measured. In assessing a board's evaluation policy we need to look at how wide a range of learning outcomes they are willing to be accountable for.

Another choice of an achievement target in the evaluation of a board's "benchmark" system is how the articulated learning objectives are assessed or evaluated and to what degree or level the articulated learning objectives have been

met. Standardized testing does not necessarily do this. All of the benchmark programs previously mentioned reiterate the fact that student achievement on learning outcomes cannot be assessed by standardized testing or certainly not solely by standardized testing. Standardized tests appear to be an easy way to be sure that there are board wide standards but all boards realize that these tests can only access a small number of learning objectives. The difficult task is to develop the expertise in the more subjective areas of evaluation. It would seem to me that a board that could be seen as operating at a high level would be one that was developing teacher expertise in evaluation. Educators are quite adept at administering paper and pencil tests to assess basic skill levels. Where educators need guidance is in developing assessment strategies for such learning outcomes as recognizing and using patterns, working collaboratively, or developing critical thinking. Some of the boards are meeting these needs better than others. The Toronto Board stated that they wanted to be sure that they did not address accountability by using standardized testing. They wanted to match the assessment of individual students with the broad spectrum of learning outcomes presented by the Ministry and Board guidelines. (Larter, 1991) They wanted to look at how students were performing in relation to other students with a holistic view of the child as learner. The Frontenac board is using samples of holistic scored student writing to help

individual teachers in their assessment of thier own student's writing. Many boards are willing to display faith in what is happening in their classrooms, others show a mistrust of what is happening in their schools. A high level "benchmark" system will help the teacher grow and develop professionally in the area of evaluation. A wide variety of assessment tools should be given to teachers as well as a plan to develop the teachers' understanding of those assessment tools. Thus the system will have faith that teachers do know their students and are the most capable to give a valid and reliable assessment of their students.

These were two achievement targets that we could use to assess a board's system of evaluation or "benchmarks". There are many other achievement targets that could be considered. The rigor or reliability and validity of each system could be benchmarked or the system of reporting evaluation to parents could be benchmarked. We also need to understand that a benchmarking system implies growth. Everyone of the benchmarking systems mentioned would describe their system as a model of growth. Students will all reach the benchmark but they may reach it at different times. It is a continuum. And so it is with the boards' benchmarking systems. Some are in their developmental stages while others are at the implementation phase. Just as educators don't assess a student before having an adequate amount of information, it is probably unfair to assess many of these benchmark systems

at this point. But it is important to listen to the discussion and to try to make some sense of it.

In attempting to make some sense out of the "benchmark" issue, each board has had to look around to see what other boards are doing. This naturally brings about collaboration and dialogue. School boards across Ontario are talking to one another. Collaboration is necessary for many boards. The Toronto Board of Education Benchmark project and the Ottawa Learning Profiles Project are extremely expensive ventures. Only a very large board could afford such an undertaking. In looking to one another for support, each board across the province is broadening its understanding of the evaluation issue and benefitting from being part of the discussion.

Selected References

Clark, John. (December, 1990). "Benchmarks in Mathematics". Ontario Mathematics Gazette v.29:2. p. 17-19.

Jones, Ann, Brown, F. and Stalkie, B. (March 1992). "The Ottawa Benchmark Project". Presentation at Ontario Association for Curriculum Development Conference.

Larter, Sylvia. (1991). Benchmarks - The development of a New Approach to Student Evaluation. Toronto: Toronto Board of Education.

Larter, Sylvia. (September, 1991). Benchmarks - One board's response to the testing problem. Federation of Women Teachers Association of Ontario Newsletter. p. 2-9.

Ministry of Education, Ontario. (1991). Benchmarks and Evaluation: A Provincial Perspective. Toronto: Ministry of Education.

Ministry of Education, Ontario. (February, 1992). Measuring Success: Educational Accountability in Ontario. Toronto: Ministry of Education.

North York Board of Education. (February, 1991). The Benchmark Vol 1, No. 1. Toronto: North York Board of Education.

North York Board of Education. (1992) Student Outcomes in Reading and Writing: P/J and Transition Years. Toronto: North York Board of Education.

Ottawa Board of Education. (1991). Benchmarks - Student Achievement Standards. Ottawa: Ottawa Board of Education.

Toronto Board of Education. (1991). Benchmark Libraries. Toronto: Toronto Board of Education.

Toronto Board of Education. (May 1991). Fact Sheet #6, Benchmarks. Toronto: Toronto Board of Education.

T.V.O. (February, 1992) Benchmarks Teleconference, (video). Toronto: T.V.O.

Winks. (1990). "The Learning Profile Project". Learning. Ottawa: Ottawa Board of Education.